

## Tilburg University

### Identifying the important factors in simulation models with many factors

Bettonvil, B.W.M.; Kleijnen, J.P.C.

*Publication date:*  
1991

[Link to publication in Tilburg University Research Portal](#)

*Citation for published version (APA):*

Bettonvil, B. W. M., & Kleijnen, J. P. C. (1991). *Identifying the important factors in simulation models with many factors*. (Research memorandum / Tilburg University, Department of Economics; Vol. FEW 498). Unknown Publisher.

#### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

#### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

CBM

CBM  
R

7626  
1991  
498

SITY

TEIT

RABANT

POSTBOX 90153

5000 LE TILBURG

THE NETHERLANDS



DEPARTMENT OF ECONOMICS  
RESEARCH MEMORANDUM





**IDENTIFYING THE IMPORTANT FACTORS IN  
SIMULATION MODELS WITH MANY FACTORS**

Bert Bettonvil  
Jack P.C. Kleijnen

**FEW 498**

*150*  
*330.115.73*



# IDENTIFYING THE IMPORTANT FACTORS IN SIMULATION MODELS WITH MANY FACTORS

BERT BETTONVIL AND JACK P. C. KLEIJNEN

Katholieke Universiteit Brabant (Tilburg University), Tilburg, The Netherlands

In practice, simulation models usually have a great many parameters and input variables. This paper presents a screening technique, which identifies the really important factors. The technique treats the simulation model as a black box and uses a regression metamodel to approximate the input/output behaviour of that black box. The metamodel can account for fitting errors with unknown variance and for interactions among factors. The technique requires relatively few simulation runs. It applies to both random and deterministic simulations. The technique is demonstrated through a case study of a complicated ecological model.

(SENSITIVITY ANALYSIS; WHAT-IF; SELECTION AND RANKING;  
REGRESSION ANALYSIS; EXPERIMENTAL DESIGN)

## 1. Introduction

In *real-life* experiments it is impossible to control hundreds of factors, but in *simulation* it is hardly a problem to experiment with so many factors. For example, in a queuing network there are many 'experimental factors': parameters (such as service rates), input variables (say, number of servers), and behavioural relationships (for example, queuing discipline). In this paper we present an example of a deterministic simulation model, namely a model of the effects of carbondioxide or CO<sub>2</sub> and other gases on the global temperature; we determine the importance of as many as 281 factors in that model. Our technique applies to both *random* and *deterministic* simulations. We wish to determine the global effects, not the local (marginal) factor effects.

The principle of *parsimony* implies that to explain the behaviour of a system, scientists should use as few factors as possible; that is, they should not state 'everything depends on everything else'. Consequently we assume that only a few factors are really important, even though there are a great many factors in the simulation experiment. *Screening* is the process of searching for the important factors among a great

many potentially important factors. In the initial phase of most simulation studies, screening should take place. Unfortunately, many users of simulation intuitively select a small number of factors and experiment with these factors only. This practice severely limits the generality of the conclusions. The screening process results in a list of important factors, which can be further explored in later experiments. Assuming the metamodel is correct, our technique also gives the magnitudes of the individual important factor effects. In a follow-up experiment the input-output behaviour can be studied more precisely; this is not part of our paper.

We treat the simulation model as a *black box*, whereas techniques such as Perturbation Analysis (see Ho et al., 1984) 'open' the box. We approximate the input-output behaviour of the simulation model through a *regression metamodel* with main effects and -where necessary- interactions. The black box approach has advantages and disadvantages. The advantage is that the approach is simple and robust. The disadvantage is that the approach cannot exploit the special structure of the simulation model at hand. Other techniques, for example group screening, also use regression metamodels, but these techniques are less 'efficient'; that is, they require more simulation runs, as we shall see.

This paper is organized as follows. In § 2 the basics of our technique are explained, assuming the simplest metamodel, namely a model that fits the simulation model perfectly. In § 3 this -unrealistic- metamodel is replaced by a model with fitting errors with known variance. In § 4 we deal with unknown variance. In § 5 the metamodel is augmented with interactions among factors. In § 6 the technique is applied to the complicated ecological simulation model mentioned above. For details we refer to Bettonvil (1990).

## 2. Basics of the procedure

### 2.1 Mathematical representation of simulation models

A simulation model can be represented mathematically as

$$y = s(v_1, \dots, v_j, \dots, v_N, r), \quad (1)$$

where  $s(\ )$  denotes the mathematical function that is specified by the computer simulation program;  $v_j$  denotes the  $j$ th factor or input of that program with  $j = 1, 2, \dots, N$ ;  $N$  is a known positive integer;  $r$  is the pseudorandom number seed; and  $y$  is the simulation output. We elaborate this representation as follows.

The simulation users are supposed to know their *domain of interest*; that is, they know the ranges of the quantitative factors:

$$L_j \leq v_j \leq H_j, \quad (2)$$

where  $L_j$  and  $H_j$  denote the lower and upper value of quantitative factor  $j$  (qualitative factors will be discussed later). We assume that the area of interest is an  $N$ -dimensional rectangle. Note that the Euclidian distance from a corner of that rectangle to the origin increases with  $N$ . The *output* of a simulation model consists of a time series; for example, waiting times of successive customers or temperature increases in successive years. We assume that this time series is characterized by a single number, say, average waiting time or total temperature increase over a fixed period.

A fundamental assumption that we shall use throughout this paper is that the user knows the *direction* of the influence a factor has on the output  $y$ , if that factor has any effect at all. We define  $L_j^*$  as the value of factor  $j$  that generates a low value for the output  $y$ ;  $H_j^*$  is the value of factor  $j$  that generates a higher value for  $y$ , provided this factor has any effect at all. For example, in queuing simulations the user expects that the number of servers reduces the mean waiting time; hence  $L^*$  of the original variable corresponds with  $H$  in (2).

It is convenient to transform the original quantitative variables  $v$  linearly into *standardized* variables  $x$  that vary between 0 and 1:

$$x_j = a_j v_j + b_j \quad (3)$$

with

$$a_j = 1/(H_j^* - L_j^*) \text{ and } b_j = -L_j^*/(H_j^* - L_j^*). \quad (4)$$

Qualitative variables (such as priority rule) are represented by standardized variables that have only two values, 0 and 1, where 0 corresponds with the level that generates a lower output.

Note that the *seed*  $r$  plays no role in deterministic simulation. In random simulation, each process (for example, arrival and service process respectively) needs its own seed if variance reduction techniques such as common and antithetic pseudorandom numbers are applied. These seeds can be selected by calling the computer's internal clock. Then  $r$  denotes a vector of seeds.

## 2.2 Simplest metamodel

The simplest metamodel has first-order or main effects  $\beta_j$ , besides the overall mean  $\beta_0$ , and fits *perfectly*:

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_N x_N. \quad (5)$$

The assumption of perfect fit holds only if the simulation model (1) is exactly given by (3) and (5); i.e., the simulation and the metamodel are identical. We use this metamodel solely to explain the basics of



our technique; in § 3 we shall introduce fitting errors. The assumption of known directions of effects implies that all main effects are non-negative:  $\beta_j \geq 0$ ; in this section positive effects are called important. To determine the effects in a first-order metamodel, it suffices to consider only two levels per factor. We take the two extreme values for each factor:  $x_j$  is 0 or 1. Our technique is *sequential*; that is, the factor combinations to be simulated depend on the results of previous combinations simulated. In other words, the experiment consists of stages. First we explain the technique informally.

Let  $y_{(j)}$  denote the simulation output observed when the factors #1 through #j are switched on and the remaining factors are off; (5) yields

$$y_{(j)} = \beta_0 + \beta_1 + \dots + \beta_j \quad (j = 0, 1, \dots, N). \quad (6)$$

Consequently  $y_{(j)}$  is a nondecreasing sequence as  $j$  increases. In stage #0 we obtain the two extreme output observations  $y_{(0)}$  and  $y_{(N)}$ : we run the simulation program once with all factors 'turned off'; that is, all factors are at their low values ( $v_j = L_j^*$  or  $x_j = 0$ ) and once with all factors turned on ( $x_j = 1$ ). Substitution into (6) yields

$$y_{(0)} = \beta_0 \text{ and } y_{(N)} = \beta_0 + \beta_1 + \dots + \beta_j + \dots + \beta_N. \quad (7)$$

If and only if (iff) all factors are unimportant, these two outputs are equal. We point out that the main effects cannot compensate each other as we assumed non-negative effects. So if no factor is important, then after just two runs we stop. Suppose that at least one factor is important. Then (7) implies that  $y_{(N)} > y_{(0)}$  and we must run more simulations. We call our technique *bifurcation* because the set of 'remaining' factors is split into two parts (after stage #0, if  $y_{(N)} > y_{(0)}$ , all factors remain to be investigated). If the number of factors is a power of two ( $N = 2^m$  with positive integer  $m$ ), then we split the number of remaining factors into two halves (in § 2.4 we shall return to this assumption, which will turn out not to be essential).

In stage #1 we run the simulation with the first half of all the remaining factors turned on; that is,  $x_1$  through  $x_{N/2}$  have the value 1. This yields

$$y_{(N/2)} = \beta_0 + \beta_1 + \dots + \beta_{N/2}. \quad (8)$$

Equations (7) and (8) show that iff the first half of the factors contains at least one important factor, then  $y_{(N/2)} > y_{(0)}$ . Iff the second half contains at least one important factor, then  $y_{(N/2)} < y_{(N)}$ . So in stage #1 just one simulation output is required, and this output is compared to the two outputs of the preceding stage. It is convenient to introduce a specific example to further discuss our technique. We use an example of Jacoby and Harrison (1962) with  $2^7$  factors and only the factors #68, #113, and #120 are important, and let  $\beta_{1,128}$  denote the sum of  $\beta_1$  through  $\beta_{128}$ , and so on; also see Figure 1. In (8) we get

$N/2 = 2^6 = 64$ , and  $y_{(64)} = y_{(0)}$  but  $y_{(64)} < y_{(128)}$ . Hence we can *eliminate* all factors in the first half; that is, after only three runs we know that the factors 1 through 64 have no effects and that there is at least one important factor in the second half of the group of factors.

INSERT FIGURE 1

In stage #2 we concentrate on the remaining factors (labelled 65 through 128). Again we split those factors into two (sub)groups of equal size. The first group consists of the factors #65 through #96; the other group comprises the factors #97 through #128. In the next run we switch on the factors of the first group, while all factors in the preceding groups are also switched on:

$$y_4 = y_3 + \beta_{(N/2 + 1)} + \dots + \beta_{(N/2 + N/4)}, \quad (9)$$

where  $N/2 + 1 = 65$  and  $N/2 + N/4 = 3N/4 = 96$ . In the example the important factor #68 makes  $y_4$  larger than  $y_3$ , and the important factors #113 and #120 make  $y_4$  smaller than  $y_2$ . So at stage #2 we obtain a single observation and compare it with the two relevant observations of the preceding stages. In the example no factors can be eliminated at this stage as the important factors are not *clustered*.

In stage #3 we must investigate two groups of factors (as no factors were eliminated at the preceding stage). Each group is split into half. And so the bifurcation continues. Figure 1 shows that after only 16 observations the three important factors are identified and their individual main effects are computed. For example, in the last stage three outputs are obtained; in these runs the first 67, 113, and 119 factors are switched on respectively, and these runs are used to compute the effects of the three important individual factors (#68, #113, and #120).

More generally, in stage #0 we always observe  $y_{(0)}$  and  $y_{(N)}$ . If these two outputs are equal, then no factor is important, and we stop. Otherwise we proceed to stage #1. In that stage we split the group of remaining (non-eliminated) factors into half, which gives two groups. We observe the simulated system with the factors in the first group switched on:  $y_{(N/2)}$ . In general, in stage #s there are  $2^s$  groups, each consisting of  $2^{m-s}$  individual factors ( $N = 2^m$ ). Suppose that  $y_{(j1)} < y_{(j2)}$  with  $j1 < j2$ , then the corresponding group contains at least one important factor. Then we proceed to the next stage and split that group into two equal parts. We switch on factors, starting with #1 through #j3 with  $j3 = (j1 + j2)/2$  (so j3 is the last factor in the newest subgroup); the other factors are off. Then  $y_{(j3)}$  is compared with  $y_{(j1)}$  and  $y_{(j2)}$ . That comparison leads to the elimination of at most one group (obviously it is impossible that both groups are eliminated). Non-eliminated groups are further investigated by sequential bifurcation. Finally, after m stages, individual factors are reached. A formal description of sequential bifurcation is given in Bettonvil (1990, pp.13 - 22).

### 2.3 Efficiency of Sequential Bifurcation Relative to Other Techniques

We quantify the relative efficiency of our technique as follows. We compute the number of simulation runs needed to find the (say)  $k$  important factors among the  $N = 2^m$  factors. Let us consider several values of  $k$ . If  $k$  is zero, then the two observations of stage #0 suffice, as we have already seen below (7). If a single factor is important ( $k = 1$ ), then we have to go through so many stages that we finally reach the stage that generates observations for individual factors. At each stage, one group is found to be important and one group is eliminated. So a single observation at each stage suffices; stage #0 is special since two outputs are observed ( $y_{(0)}$  and  $y_{(N)}$ ). Some thought shows that because there are  $2^m$  factors, the total number of runs is  $2 + m$ . Next we consider a system with two important factors ( $k = 2$ ). After stage #0 we proceed to stage #1 (and observe  $y_{(N/2)}$ ). The two important factors may happen to be *clustered*; that is, they belong to the same group. Then the other group is eliminated ( $N/2$  factors are eliminated after stage #1). However, if we have bad luck, the two important factors are in different groups and at each stage two outputs must be generated, namely one in each group. The total number of runs is then  $3 + 2(m-1) = 1 + 2m$ . Bettonvil (1990, pp. 24-25) proves that in the *worst case* (important factors are not clustered) the number of runs (say)  $n$  is given by the following equation if the number of important factors ( $k$ ) is a power of two; otherwise this equation gives an approximation:

$$n = 1 + k \{^2\log (2N/k)\}. \quad (10)$$

In the Jacoby and Harrison example,  $N$  is 128 and  $k$  is 3, so the approximation gives  $n = 20.2$ . Because the important factors showed some clustering, the actual number of runs is 16.

We compare the number of runs for our technique with the number required by other techniques that use regression analysis and experimental design to screen simulation models treated as black boxes. Mauro and Burns (1984) compare random and two-stage group-screening designs, and conclude that *random* designs are inefficient. *Two-stage group-screening* was introduced by Watson (1961) and is evaluated by Mauro (1984). In the worst case the number of runs- given an optimal group size- is

$$n_2 = 2(kN)^{1/2}. \quad (11)$$

Comparison with (10) proves that our technique is more efficient; Table 1 gives numerical results.

Patel (1962) and Li (1962) generalize two-stage group-screening to *multi-stage group-screening*, assuming that each factor has an a priori probability (say)  $p$  of being important. This probability yields an optimal group size and an optimal number of stages, which result in an optimal number of runs:

$$n_3 = 1 - N p e \ln p \approx k e \ln N. \quad (12)$$



Comparison with (10) proves that our technique is still more efficient; Table 1 gives numerical results.

Recently Morris (1987) developed *multiple grouping*, which is a variant of two-stage group-screening. Bettonvil (1990, p.29) gives the (complicated) formula for the number of runs; Table 1 gives results for a specific case.

*Jacoby and Harrison (1962)'s sequential bifurcation* uses two observations where we use a single one; that is, they use a less efficient design at each stage.

Note that Bettonvil (1988) uses a design that switches on only the factors *within* the group to be investigated, whereas we switch on all factors starting with factor #1 and ending with the last factor within the group. The former design requires the same number of runs but complicates the analysis when random fitting errors are assumed.

Note further that *search linear* models were introduced by Srivastava (1975), assuming that the number of important factors is small but *known*. The author gives specific designs only for  $k$  equal to one or two. Ghosh (1979) generalizes this approach to multi-stage designs in such a way that his technique resembles ours. His number of runs equals ours. Because his technique uses more restrictive assumptions than ours, it will not be further considered in this paper.

Table 1 shows the number of runs required by the various techniques in the worst case if the number of factors is 1024 ( $m = 10$ ) and the number of important factors ranges between zero and eight. For the first three techniques we assume that the a priori probability  $p$  is known exactly ( $p = k/N$ ), which is an assumption that favours those techniques. The results of Table 1 clearly show the efficiency of our technique.

#### INSERT TABLE 1

Using the a priori probability  $p$  and assuming  $N = 2^m$ , Bettonvil (1990, pp. 34-35) derived the *expected* number of runs:

$$E(n) = 1 + N - \sum_{j=1}^m 2^{m-j} (1-p)^{2^j} \quad (13)$$

Morris (1987) gives the expected number of runs for his technique and for two-stage group-screening. Table 2 shows that our technique is more efficient under this criterion, too.

#### INSERT TABLE 2

## 2.4 Relaxing Some Assumptions

We now assume that the number of factors ( $N$ ) is not necessarily a *power of two*. It is simple to adapt our technique: we add dummy factors after the last factor so that the new set of factors consists of  $2^m$  factors. In this way the unimportant dummy factors are already clustered; clustering reduces the maximum number of runs, as we saw in § 2.2. For example, if there are twelve factors, then we do not split them into two groups of six each but into one group of eight and one group of four factors. Bettonvil (1990, pp.40-43) shows that both the expected and the maximum number of runs are smaller in the latter type of splitting. The clustering argument implies that our technique is most efficient if the factors are labelled from 1 through  $N$  in *increasing* order of importance. Then the important factors are clustered in the (second) group of four factors, not in the (first) group of eight factors. So our technique can take advantage of a priori knowledge about factor effects.

So far we assumed that a factor is *important* if its first-order effect is positive, not zero. Practitioners will declare a factor important only if it has a 'sizable' (positive) effect. Our technique does not require the practitioner to specify a critical value (say)  $\delta$  that has to be exceeded by a main effect in order to be declared important. The technique implies that a group of factors that seems important at the end of a stage is split into two smaller subgroups; in the next stage those two subgroups are investigated (all factors from factor #1 through the last factor of the first subgroup are switched on). The new run splits the sum of main effects of the factors in the original group into two sums: sum #1 (and sum #2) equals the sum of the main effects of the factors in the first subgroup (and the second subgroup respectively). Obviously no individual main effect can exceed the sum of which it is a part. As soon as the user considers this sum to be 'small', the investigation of this subgroup is stopped. For example, suppose there are 24 factors and that their main effects have the magnitudes shown by the vertical solid bars in Figure 2. In stage #0 (not displayed) we observe  $y_{(0)} = 0.0$  and  $y_{(N)} = 2388.2$  so the upper limit for the sum of individual effects after two runs is (say)  $U_2 = 2388.2 - 0.0 = 2388.2$ . We would stop if we were interested only in effects exceeding that limit. Suppose that limit is too coarse. As we proceed through various stages, these upper limits decrease. The dotted horizontal lines in Figure 2 are the upper limits after 11 through 17 runs. These limits decrease from 383.6 to 139.7 as we go from 11 to 17 runs. So after 13 runs we know that the factors #17 and #20 are 'important'; that is, they have main effects larger than 217.9. After 17 runs we have identified the eight most important factors (#14, ..., #24), and we know that the remaining factors have no main effects larger than 139.7. Notice that this example does not favour our technique: the clustering is bad.

INSERT FIGURE 2.



### 3. Random Fitting Errors

Now we drop the assumption of perfect fit, which we made in (5). So we assume that the metamodel has fitting errors (say)  $\epsilon$ . We further assume that these errors are additive and distributed normally. The normal distribution is assumed for both random and deterministic simulation models, which are approximated by metamodels. The modelling of fitting errors in *deterministic* simulation is a controversial issue, discussed at length by Kleijnen (1990). We assume that in deterministic simulation the fitting errors may be modeled as a normal variable with expectation zero iff the metamodel is adequate, and with a constant variance (say)  $\sigma^2$ . Whereas Sacks, Welch, Mitchell and Wynn (1989) assume a covariance stationary process, we assume *white noise*; that is,  $\epsilon$  is independently distributed. In summary, we assume that  $\epsilon$  is Normally, Independently and Identically Distributed (NIID) with zero mean and variance  $\sigma^2$ :

$$\epsilon \sim \text{NIID} (0, \sigma^2) \quad (14)$$

For *random* simulation we assume in this paper that  $\epsilon$  also satisfies (14). (In reality these fitting errors may have variances that change as the inputs change; if common or antithetic pseudorandom seeds are used, then these errors are no longer independent.) Initially we shall assume that the variance  $\sigma^2$  is known; next we shall assume an unknown variance. We shall also investigate the sensitivity of our technique to nonnormality.

Because there is noise in the metamodel, we must adjust our definition of *importance*. We want our technique to detect the factors with first-order effects that are 'large' relative to the noise; that is, we wish to select main effects larger than a prespecified multiple  $\delta$  of the standard deviation  $\sigma$ :

$$\beta_j \geq \delta \sigma. \quad (15)$$

We wish to detect such factors with a prespecified probability  $P^*$ . This probability may be compared to the complement of the type-I error in hypothesis testing. We also desire that the probability of declaring unimportant factors to be important is 'small'. Finally we want the number of simulation runs to be small.

#### 3.1 Known variance

Beatonvil (1990, pp. 54-88) derives three statistical techniques that handle white noise with known variance. One technique resembles the approach followed for perfect fit, but is inferior when compared with the other two techniques. The second technique uses Bechhofer's (1954) table with critical values for selecting the  $t$  populations with the highest means from  $k$  normal populations with known common

variance  $\sigma^2$ , given a prespecified probability of correct selection  $P^*$ . The third technique uses *sums of squares*. It gives results that are no worse; moreover it can best be adapted to unknown variances. We now present the third technique.

In § 2.3 we saw that if there are no fitting errors, then our sequential bifurcation requires  $m+2$  runs to detect an important factor. Each factor has its own path, always starting with the two outputs of stage #0 but finishing with the  $(m+2)th$  run that identifies a specific factor. To identify factor  $j$ , one run is needed with the factors #1 through #(j-1) switched on and one run with the factors #1 through #j on. Now there are fitting errors and the  $m+2$  outputs that are used to detect factor  $j$  are denoted by  $y_{(t)}^j$  ( $t=0,1,\dots,m+1$ ), where  $y_{(0)}^j = y_{(0)}$  and  $y_{(m+1)}^j = y_{(N)}$ ; the order of expected values is known since switching factors on cannot decrease expectations; see (9)):

$$E(y_{(0)}^j) \leq \dots \leq E(y_{(t)}^j) \leq \dots \leq E(y_{(m+1)}^j) \quad (t=0,1,\dots,m+1). \quad (16)$$

We further define  $\hat{y}_{(t)}^j$  ( $t=0,1,\dots,m+1$ ) as estimates that correspond with (16) and with the definition of importance in (15); that is, these estimates are required to be non-decreasing and show a sizable gap as soon as factor  $j$  is switched on:

$$\hat{y}_{(0)}^j \leq \hat{y}_{(1)}^j \leq \dots \leq \hat{y}_{(g)}^j \leq \hat{y}_{(g+1)}^j - \delta \sigma \leq \dots \leq \hat{y}_{(m+1)}^j - \delta \sigma, \quad (17)$$

where  $y_{(g)}^j = y_{(j-1)}$  and  $y_{(g+1)}^j = y_{(j)}$ . Moreover these estimates must minimize the sum of squares formed by the observed and the estimated outputs:

$$SS = \sum_{t=0}^{m+1} (Y_{(t)}^j - \hat{Y}_{(t)}^j)^2. \quad (18)$$

Suppose the observed outputs are non-decreasing and show the desired gap; see (17) with carets deleted. Then (18) is minimized subject to (17) if we take the observed and the estimated outputs equal; the minimal SS, denoted as  $SS_{\min}$ , becomes zero. So if  $SS_{\min}$  is small, then the factor is important. Bettonvil (1990, pp.74-75, 208-214) uses intuition, supported by Monte Carlo experimentation, to decide that  $SS_{\min}/\sigma^2$  should be treated as a chi-square statistic with  $m+1$  degrees of freedom:

$$SS_{\min}/\sigma^2 = \chi_{m+1}^2. \quad (19)$$

This statistic should be tested at the level  $1 - 2\alpha$  with  $\alpha = 1 - P^*$ . At each stage of the sequential bifurcation the minimal sum of squares is computed for the observations  $y_{(t)}^j$  obtained so far. These minima increase as more observations become available. Hence as soon as a group of factors yields a significantly large statistic, that group is eliminated. (The algorithm for solving this minimization problem is not presented because it is trivial and does not apply for situations with unknown variances.)

We present some Monte Carlo results. Suppose there are 256 factors ( $N = 2^8$ ). Bettonvil (1990, p.60) investigates five situations: (i) no factors important; (ii)  $\beta_1 = \delta\sigma$  and all other factors unimportant; (iii)  $\beta_{86} = \delta\sigma$  and all other effects zero; (iv)  $\beta_{241} = \delta\sigma$  and all other effects zero; and (v) factors #1, #86, and #241 have effects equal to  $\delta\sigma$  and the other factors have no effects. Obviously, if factors had effects larger than  $\delta\sigma$ , then it would be easier for our technique to detect these effects. The labels of the important factors are selected such that various paths are followed: an important factor #1 means that the left branch is always followed; #86 means that we go left, right, left, right, and so on; #241 implies that we first go right four times, then left four times. Suppose further that  $\alpha (= 1 - P^*)$  equals .05, .005, and .0005;  $\delta$  equals 10, 8, and 6. To obtain reliable Monte Carlo estimates, the experiment is repeated 1000 times. The results are the estimated probability of detecting the truly important factors, the average number of unimportant factors declared (falsely) to be important, and the average number of runs needed by sequential bifurcation. The Monte Carlo experiment uses common random numbers, so that results can be better compared. Table 3 demonstrates that if the user wants to 'play safe' ( $\delta$  small or  $P^*$  large), then the number of unimportant factors declared important and the number of runs increases. As  $\delta$  changes, the estimated realized  $\alpha$  remains constant. Many more Monte Carlo results are given by Bettonvil (1990).

Note that the pseudorandom numbers are obtained through the NAG generator, which is multiplicative with multiplier  $13^{13}$  and modulus  $2^{59}$ ; we use the seed 2016171. The Gaussian variables are generated through the well-known Box-Muller transformation.

### INSERT TABLE 3

Note that if the number of factors is not necessarily a power of two, then a simple adaptation of our technique adds dummy factors so that the total number of factors becomes a power of two again. Bettonvil (1990, pp.83-85, 191-199) derives a different solution that may save runs and may decrease the number of falsely detected (unimportant) factors; the probability of detecting important factors is not affected.

### 3.2 Robustness of Sequential Bifurcation

How sensitive is the technique presented in the preceding subsection to the assumptions of non-negative effects only and to nonnormality? To answer the first question we use a Monte Carlo experiment with parameters taken from Mauro and Burns (1984), which presumably do not favour our technique. This experiment has 100 factors with 20 *negative* effects, 32 positive effects, and 48 zero effects; these effects range between -3.85 and 9.57;  $\sigma = 1$ . The results in Bettonvil (1990, pp.89-92) demonstrate that effects equal to  $\delta$  are detected with a smaller probability than prescribed; of course as the magnitude of an effect increases, so does the probability of detection. The technique based on



Bechhofer's statistic (see § 3.1) gives probabilities that may exceed the prescribed probability when the effect exceeds  $\delta$ .

Bettonvil (1990, pp.93-108) investigates *nonnormality* through a Monte Carlo experiment that includes both symmetry and kurtosis: distributions of the uniform, exponential, and double-exponential (Laplace) type are studied. The results demonstrate that heavy tails give outliers, which make the technique miss important factors more often than prespecified through  $P^*$ .

The robustness study indicates that Bechhofer's statistic gives a slightly more robust technique. However, the sums of squares technique can be better extended to the realistic case of unknown variances, as we shall see next.

#### 4. Unknown Variance

In selection and ranking problems the unknown variance is usually estimated from a pilot sample; the critical statistic then depends on the number of pilot runs and requires the numerical solution of a complicated integral; see Dudewicz and Koo (1982). We follow a completely different approach.

In sequential bifurcation the observed outputs reveal both the effects of important factors (these factors cause a 'gap' in the sorted outputs) and the variance (unimportant factors give outputs that differ only because of noise). We can compute  $SS_{\min}$  through (17) and (18), assuming some value for the variance (say  $\sigma_0^2$ ). If a group of factors is unimportant for *any positive value of the variance*, then this group is eliminated. As long as there are positive values of the variance that yield a minimum sum of squares smaller than the critical value of the chi-square statistic, we continue the bifurcation process. Bettonvil (1990, pp.126-127) gives detailed numerical examples. Our technique assumes that a group of factors is unimportant whenever switching that group on yields an output *smaller* than leaving it switched off, given fixed values for the other factors. Bettonvil (1990, pp. 129-130) proves that this assumption implies

$$\delta^2 \geq 2 \chi^2, \quad (20)$$

where  $\chi^2$  stands for the  $1 - 2(1 - P^*)$  quantile of the chi-square statistic with  $m + 1$  degrees of freedom; also see (19). If there is no positive value for the variance that gives a significantly small minimum sum of squares, then we eliminate the corresponding group of factors; otherwise we take the next observation 'down the bifurcation tree'. A property of the technique is that when we take a new observation, the range of possible values of the variance does not increase: information pays off. Finally we may arrive at the individual factor (after  $m+2$  runs) and still have a range of possible values for the variance. Then we conclude that this factor is indeed important. Details of the computational procedure are given by Bettonvil (1990, pp. 125-134, 215-219); it boils down to solving quadratic polynomials. Monte Carlo

results are given in Table 4, the analogue of Table 3, except that now the variance is unknown.

#### INSERT TABLE 4

Comparison with Table 3 shows that  $\delta = 6$  is not combined with  $1 - P^*$  equal to .005 or .0005, because of (20). Table 4 demonstrates that the important factors are found more frequently than is specified by  $P^*$ ; more than 99% of the unimportant factors are recognized as such; the average number of runs increases by roughly 50% when the variance is unknown. We find these results quite satisfactory.

We examine the sensitivity of the technique to *negative* effects. We use a Monte Carlo experiment with the same parameters as in § 3.2. Comparison with that subsection shows that an unknown variance increases the probability of detecting important factors. That probability is still smaller than  $P^*$  if  $\beta_j = \delta \sigma$  since a negative factor effect is interpreted as noise: the variance is overestimated and the important factor is underestimated.

*Nonnormality* may mean high kurtosis and skewness. Heavy tails do not have much effect, whereas asymmetry demands caution, as the results in Bettonvil (1990, pp.139-140) demonstrate.

### 5. Interactions

We can easily extend sequential bifurcation to metamodels either with only *first-order* interactions or with *all* interactions present. The latter metamodel seems less interesting; for completeness' sake it is presented in Bettonvil (1990, pp. 220-227).

#### 5.1 Perfect Fit

The metamodel with first-order interactions and without fitting errors extends (5) to

$$y = \beta_0 + \sum \beta_j x_j + \beta_{12} x_1 x_2 + \dots + \beta_{N-1,N} x_{N-1} x_N. \quad (21)$$

We shall see that it is convenient to use a different parametrization: replace  $x_j$  by  $z_j$  where  $z_j$  is -1 iff  $x_j = 0$ , and  $z_j = 1$  iff  $x_j = 1$ . Then (21) is equivalent to

$$y = \gamma_0 + \sum \gamma_j z_j + \gamma_{12} z_1 z_2 + \dots + \gamma_{N-1,N} z_{N-1} z_N. \quad (22)$$

Equations (21) and (22) yield

$$\gamma_{jj'} = \beta_{jj'}/4, \quad (23)$$

$$\gamma_j = \beta_j/2 + \sum_{j' \neq j} \beta_{jj'}/4, \quad (24)$$

$$\gamma_0 = \beta_0 + \sum_{j=1}^N \beta_j/2 + \sum_{j=1}^{N-1} \sum_{j'=j+1}^N \beta_{jj'}/4 \quad (25)$$

Initially we again suppose that the number of factors is a power of two ( $N = 2^m$ ). In the metamodel with main effects only we assumed that these effects were non-negative. Now we assume that if a factor is switched on while the other factors are kept constant, then the output does not decrease. This *monotonicity* assumption means that

$$y(x_1, \dots, x_{j-1}, 0, x_{j+1}, \dots, x_N) \leq y(\dots, 1, \dots) \quad (26)$$

holds for any  $j$  and for all values of the  $N-1$  remaining factors. We call factor  $j$  *important* in a *narrow* sense iff there is at least one combination of the remaining factors that yields an increase of output larger than a given constant  $\delta$  when factor  $j$  is switched on:

$$y(x_1, \dots, x_{j-1}, 1, x_{j+1}, \dots, x_N) > y(\dots, 0, \dots) + \delta. \quad (27)$$

Bettonvil (1990, pp.148-150) proves that (26) implies that the interactions are small relative to the main effects. Factor  $j$  is *important* in a *wide* sense if switching this factor on and averaging over all  $2^{N-1}$  combinations of the remaining  $N-1$  factors yields an increase larger than  $\delta/2$  when compared with switching that factor off and averaging over all  $2^{N-1}$  combinations. It can be proved that, as this terminology suggests, a factor that is important in the narrow sense is always important in the wide sense too. By using the definition of importance in the wide sense, we can easily extend sequential bifurcation to metamodels with interactions, as follows.

Now that there are interactions we double the number of runs: besides the runs observed for metamodels with main effects only, we also observe the *mirror* runs. So if we observed  $y_{(j)}$  (defined in (9)) we now also observe  $y_{-(j)}$ , which is defined as the output if the first  $j$  factors are switched off and the remaining factors are switched on. Hence  $y_{-(0)} = y_{(N)}$  and  $y_{-(N)} = y_{(0)}$ . Bettonvil (1990, pp.152-153) proves that the difference between the original output and its mirror output,  $y_{(j)} - y_{-(j)}$ , is a non-decreasing function of  $j$ ; and factor  $j$  is important iff

$$\{y_{(j)} - y_{-(j)}\} - \{y_{(j-1)} - y_{-(j-1)}\} > \delta. \quad (28)$$

Note that the idea of doubling the number of runs in order to identify main effects in the presence of first-order interactions resembles the 'foldover' principle presented in Box and Wilson (1951, p.35). This principle led to resolution IV designs constructed from resolution III designs. Resolution IV designs also play a role in multi-stage group-screening. See Kleijnen (1987).

Our technique for metamodels with interactions is completely analogous to sequential bifurcation for metamodels with main effects only. So in stage #0 all factors are off and all factors are on respectively, which yields  $y_{(0)}$  and  $y_{(N)}$ . Equation (28) suggests that we compute

$$\{y_{(N)} - y_{-(N)}\} - \{y_{(0)} - y_{-(0)}\} = 2y_{(N)} - 2y_{(0)}. \quad (29)$$

If the contrast in (29) exceeds  $\delta$ , we observe  $y_{(N/2)}$  and  $y_{-(N/2)}$ . We then compute

$$\{y_{(N/2)} - y_{-(N/2)}\} - \{y_{(0)} - y_{-(0)}\} \quad (30)$$

and

$$\{y_{(N)} - y_{-(N)}\} - \{y_{(N/2)} - y_{-(N/2)}\}. \quad (31)$$

If (30) does not exceed  $\delta$ , we eliminate the first half of the  $N$  factors; otherwise we proceed to the next stage. The second half is treated analogously. Finally we arrive at the individual important factors and we compute (28).

## 5.2 Fitting Errors

The metamodel with main effects only needs  $m + 2$  observations for the identification of important factors. These observations are independent and they have variance  $\sigma^2$ . Now, however, there are only  $m + 1$  independent contrasts since  $y_{(0)} - y_{-(0)} = -\{y_{(N)} - y_{-(N)}\}$ . To realize independence we can simply observe  $y_{(0)}$  and  $y_{(N)}$  twice. When we compute the sum of squares from (17) through (19), we replace the simulation outputs  $y_{(j)}$  by the contrasts  $y_{(j)} - y_{-(j)}$  and the variance  $\sigma^2$  by  $2\sigma^2$  (where  $\sigma^2$  still denotes the variance of the simulation output  $y$ ).

We can avoid the replication of the two runs in stage #0, if we assume that there is at least one important factor. Under that assumption stage #0 is always followed by stage #1, which yields  $y_{(N/2)}$  and  $y_{-(N/2)}$ . For the factors in the first half we drop  $y_{(N)} - y_{-(N)}$  and replace it by  $y_{(N/2)} - y_{-(N/2)}$ ; for the second half we drop  $y_{(0)} - y_{-(0)}$  and replace it by  $y_{(N/2)} - y_{-(N/2)}$ .

We need no new Monte Carlo experiments to estimate the performance of our technique in case of interactions, since our technique for the metamodel with interactions is completely analogous to the technique for the metamodel with main effects only. So the behaviour found in § 3.1 also holds now, except for the number of runs that is doubled; see Bettonvil (1990, pp.161-172). When the variance is unknown, we proceed as in § 4.



## 6. Case Study: Ecological Simulation Model

We apply our technique to a deterministic model of the 'greenhouse effect'. This model tries to explain how the global temperature increases as a consequence of the increase of gases like carbondioxide ( $\text{CO}_2$ ). The model has been developed by the Dutch National Institute of Public Health and Environmental Protection, abbreviated in Dutch to RIVM. They call this model IMAGE, Integrated Model to Assess the Greenhouse Effect. IMAGE consists of many modules or submodels, as Figure 3 shows.

INSERT FIGURE 3

An essential part of IMAGE represents the global circulation of  $\text{CO}_2$  in the atmosphere, the oceans, and the terrestrial biosphere. We apply sequential bifurcation to this part. The output  $y$  denotes the simulated  $\text{CO}_2$  concentration in the year 2100. We investigate 281 factors and specify a metamodel with first-order interactions. The system experts specify the range of these 281 factors, as required in (2). Using prior knowledge we label the factors such that their effects are, we hope, sorted in increasing order. Stage #0 gives  $y_{(0)} = 195$  and  $y_{(281)} = 4316$ . The system experts, however, find these two values unrealistic. (We change 281 factors, which yields a Euclidian distance of 16.8 from the origin of the standardized factors, whereas the system experts changed only 20 factors in earlier experiments, which gives a distance of 4.5.) Therefore we reduce the factor ranges to 1/10 of their original values. This reduction of the experimental domain yields  $y_{(0)} = 988$  and  $y_{(281)} = 1496$ , which is acceptable. Since the difference between these two outputs is considered to be large, we expect important factors to be present. So we proceed to stage #1, and observe  $y_{(256)}$  and its mirror observation  $y_{-(256)}$ . And so on.

After 77 pairs of runs, the upper limit is 2.5 (the idea of upper limits was presented in Figure 2). We have then found 35 factors with known effects  $\gamma_j$  that exceed 2.5; all other factors have effects smaller than 2.5 and are declared unimportant.

There is one new complication: some factors are *linearly dependent*; that is, their values add up to one (these factors denote fractions). We observe those linear combinations that yield the maximum and the minimum response respectively. If these two responses differ only a little, the group of linearly dependent factors is unimportant. It turns out that several groups of linearly dependent factors are indeed unimportant.

Accounting for linear dependencies, we find that only 15 of the 35 factors that seem important, are really important. Table 5 displays these factors, together with their effects  $\gamma_j$ , in decreasing order of importance, and their ranges (some ranges start from the high value in order to make the signs positive; see § 2.1). The number of runs so far is 160 (77 pairs and 3 pairs to resolve linear dependencies).



## INSERT TABLE 5

We *verify* the results of sequential bifurcation as follows. There are eight factors that the system experts expect to be important. Seven of these factors are factors also declared important by sequential bifurcation; one other factor, called DIFF, is not declared important by our technique. We now observe 16 combinations of these eight factors that form a resolution IV design (by definition a resolution IV design gives estimators of the main effects that are not biased by possible two-factor interactions; the estimators for the two-factor interactions, however, are biased). Comparison of Tables 5 and 6 shows that sequential bifurcation correctly declared the seven factors to be important; our technique also gives roughly the same point estimates of the main effects as the resolution IV design gives. The latter design shows that the factor DIFF is indeed 'unimportant', namely smaller than 2.5. Moreover Table 6 suggests that interactions are unimportant (the symbol (12) in that table denotes the estimated interaction between the factors 1 and 2; actually resolution IV means that this estimator is aliased with other two-factor interactions).

## INSERT TABLE 6.

We point out that some of the important factors detected by our technique (especially shifts from and to temperate forests), were neglected by the system analysts in their original experiments. So our results are unexpected and give new insight.

The case study outputs nearly always increase as additional factors are switched on. There are five exceptions among the 154 ( $= 2 \times 77$ ) outputs. These exceptions are explained by wrong coding of 'upper' and 'lower' levels (see § 2.1).

Note that reduction of the experimental area may mean that a first-order metamodel suffices, so only half the number of runs is needed.

## 7. Conclusions

In practice, simulation models usually have many factors, but only a few factors are really important. To detect these important factors we can apply sequential bifurcation. This technique treats the simulation model as a black box. It requires fewer simulation runs than other aggregating, group-screening techniques do. The experimental design used in sequential bifurcation depends on the regression metamodel. If that metamodel assumes first-order effects only, then the number of simulation runs is half the number used in metamodels that assume two-factor interactions. Sequential bifurcation can deal with fitting errors in the metamodel. We model these errors as white noise. Three statistical techniques can then be applied; one of them is the sum of squares technique, which uses the chi-square

statistic. Statistical techniques, however, require that at the beginning of the experiment the user specifies two parameters, namely  $P^*$  (the minimum probability of finding an important factor) and  $\delta$  (the minimum value that makes a factor effect 'important'). Fortunately, sequential bifurcation can be applied without specifying these two parameters: as simulation outputs are observed, upper limits for the factor effects become available, and the user can stop as soon as they find these limits sharp enough.<sup>1</sup>

<sup>1</sup> We thank dr. Peter Sander (Professor in the School of Industrial Engineering and Management, Eindhoven University of Technology), who -together with Jack Kleijnen- supervised the doctoral work of Bert Bettonvil. Bettonvil was financially supported jointly by the Eindhoven and Tilburg universities (Samenwerkingsorgaan Brabantse Universiteiten, SOBU). We also thank dr. Edward Dudewicz (Professor in the Department of Mathematics, Syracuse University) and dr. Shanti Gupta (Professor in the Department of Statistics, Purdue University) for their useful comments on parts of Bettonvil's dissertation.

### References

- BECHHOFFER, R.E., "A single-sample multiple decision procedure for ranking means of normal populations with known variances", *Annals of Mathematical Statistics* 25, (1954), 16-39.
- BETTONVIL, B., "Group-screening by sequential bifurcation", in H.Schellhaas *et.al.* (eds.), *Operations Research Proceedings 1987*, Springer Verlag, Berlin-Heidelberg, 1988.
- BETTONVIL, B., *Detection of Important Factors by Sequential Bifurcation*, Tilburg University Press, Tilburg, 1990.
- BOX, G.E.P. and K.B. WILSON, "On the experimental attainment of optimum conditions", *Journal of the Royal Statistical Society, Series B*, 13, (1951), 1-38.
- DUDEWICZ, E.J. and J.O. KOO, *The Complete Categorized Guide to Statistical Selection & Ranking Problems*, American Sciences Press, Inc., Columbus, Ohio, 1982.
- GHOSH, S., "On single and multistage factor screening procedures", *Journal of Combinatorics, Information & System Sciences* 4, (1979), 275-284.
- HO Y.C., R. SURI, X.R. CAO, G.W. DIEHL, J.W. DILLE, and M. ZAZANIS, "Optimization of large multiclass (non-product-form) queuing networks using perturbation analysis", *Large Scale Systems* 7, (1984), 165-180.
- JACOBY, J.E. and S. HARRISON, "Multi-variable experimentation and simulation models", *Naval Research Logistic Quarterly* 9, (1962), 121-136.
- KLEIJNEN, J. P. C., *Statistical Tools for Simulation Practitioners*, Marcel Dekker, Inc., New York, 1987.
- KLEIJNEN, J.P.C., "Statistics and deterministic simulation models: why not?", *Proceedings of the 1990 Winter Simulation Conference*, O. Balci (Ed.), 344-346.
- LI, C.H., "A sequential method for screening experimental variables", *American Statistical Association Journal* 57, (1962), 455-477.

- MAURO, C.A., "On the performance of two-stage group screening experiments", *Technometrics* 26, (1984), 255-264.
- MAURO, C.A. and K.C. BURNS, "A comparison of random balance and two-stage group screening designs: a case study", *Communications in Statistics: Theory and Methods* 13, (1984), 2625-2647.
- MORRIS, M.D., "Two-stage factor screening procedures using multiple grouping assignments", *Communications in Statistics: Theory and Methods* 16, (1987), 3051-3067.
- PATEL, M.S., "Group-screening with more than two stages", *Technometrics* 4, (1962), 209-217.
- ROTMANS, J., *IMAGE: An Integrated model to Assess the Greenhouse Effect*, Kluwer Academic Publishers, Dordrecht/Boston/London, 1990.
- SACKS, J., W.J. WELCH, T.J. MITCHELL and H.P. WYNN, "Design and analysis of computer experiments" (including Comments), *Statistical Science* 4(4), (1989), 409-435.
- SRIVASTAVA, J.N., "Designs for searching nonnegligible effects", in J.N. Srivastava (ed.) *A Survey of Statistical Design and Linear Models*, North-Holland Publishing Company, Amsterdam, (1975), 507-519.
- WATSON, G.S., "A study of the group screening method", *Technometrics* 3, (1961), 371-388.

TABLE 1

*Maximum number of runs if  $N = 1024$* 

Technique	Number of important factors (k)								
	0	1	2	3	4	5	6	7	8
Two-stage group-screening	2	68	96	116	136	152	168	180	192
Multi-stage group-screening	2	20	35	49	62	74	85	96	107
Morris group-screening	2	12	33	53	73	80	89	100	113
Jacoby & Harrison	3	21	39	55	71	85	99	113	127
Sequential Bifurcation	2	12	21	29	37	44	51	58	65

TABLE 2

*Expected number of runs for  $N = 1024$* 

Technique	A priori probability p			
	.0001	.001	.01	.1
Two-stage group-screening	21.9	64.3	198.2	521.1
Morris " "	10.3	26.4	130.3	521.1
Sequential Bifurcation	3.0	11.4	70.5	374.2



TABLE 3

*Performance of Sum of Squares Technique in Sequential Bifurcation*Factors #1, #86, and #241 Have Effects Equal to  $\delta\sigma$ 

	$\delta = 10$			$\delta = 8$			$\delta = 6$		
	$\alpha = .05$	$\alpha = .005$	$\alpha = .0005$	$\alpha = .05$	$\alpha = .005$	$\alpha = .0005$	$\alpha = .05$	$\alpha = .005$	$\alpha = .0005$
(1)	.974	.996	1.000	.974	.996	1.000	.974	.996	1.000
(2)	.988	1.000	1.000	.988	1.000	1.000	.988	1.000	1.000
(3)	.972	.996	1.000	.972	.996	1.000	.972	.996	1.000
(4)	.000	.000	.000	.003	.031	.204	.412	2.590	7.745
(5)	22.9	23.0	23.2	23.1	24.3	27.0	28.1	40.9	59.9

(1): Fraction of times  $\beta_1$  is found.(2): Fraction of times  $\beta_{86}$  is found.(3): Fraction of times  $\beta_{241}$  is found.

(4): Average number of incorrectly found parameters.

(5): Average number of observations.

TABLE 4

*Performance of Sum of Squares Technique for Unknown Variance*

	$\delta = 10$			$\delta = 8$			$\delta = 6$
	$\alpha = .05$	$\alpha = .005$	$\alpha = .0005$	$\alpha = .05$	$\alpha = .005$	$\alpha = .0005$	$\alpha = .05$
(1)	.987	.999	1.000	.987	.999	1.000	.992
(2)	.995	1.000	1.000	.995	1.000	1.000	.994
(3)	.984	.998	1.000	.982	.998	1.000	.983
(4)	.602	1.037	1.528	1.149	2.194	3.504	3.063
(5)	32.3	34.0	35.8	34.3	37.8	41.4	40.3

TABLE 5  
*Important Factors in IMAGE according to Sequential Bifurcation*

label , name	effect	range		meaning
250 CHREF(31)	12.7475	0.0	0.6	shift from temperate forest to agricultural land
246 CHREF(24)	8.3725	0.9	1.4	shift from temperate forest to grassland
19 TC2A	7.35	19.0	21.0	residence time in the (thick) cold mixed layer
237 CHREF(10)	7.0925	0.0	0.3	shift from temperate forest to open tropical forest
243 CHREF(20)	6.8	0.5	0.0	shift from human area to temperate forest
242 CHREF(19)	5.6	0.4	0.0	shift from agricultural land to temperate forest
241 CHREF(18)	5.26	0.5	0.0	shift from grassland to temperate forest
240 CHREF(16)	5.2075	0.3	0.0	shift from open tropical forest to temperate forest
281 STIM	4.8305	0.41	0.38	biotic stimulation factor
13 PRECIP	4.46	0.81	0.79	rate of precipitation of carbon in the oceans
86 CHAREF(2)	3.5525	0.11	0.095	fraction of charcoal formed upon burning of branches
239 CHREF(15)	3.4175	0.2	0.0	shift from closed tropical forest to temperate forest
22 IAREA(1)	3.4125	1282.23	1482.23	initial area of ecosystem 1 (tropical closed forest)
20 MFLOW	3.0475	2.27	2.37	circulating massflow (Gordon flow)
244 CHREF(21)	2.8625	0.2	0.0	shift from semi-desert to temperate forest

TABLE 6

*Important Factors in IMAGE according to  
Resolution IV Design*

(1) CHREF(31)	13.07	(0) mean	1224.33
(2) CHREF(24)	8.54	(12)	-0.05
(3) TC2A	7.31	(13)	0.10
(4) STIM	5.16	(14)	0.09
(5) PRECIP	4.42	(15)	0.08
(6) CHAREF(2)	3.54	(16)	0.07
(7) MFLOW	3.10	(17)	0.07
(8) DIFF	1.98	(18)	0.06

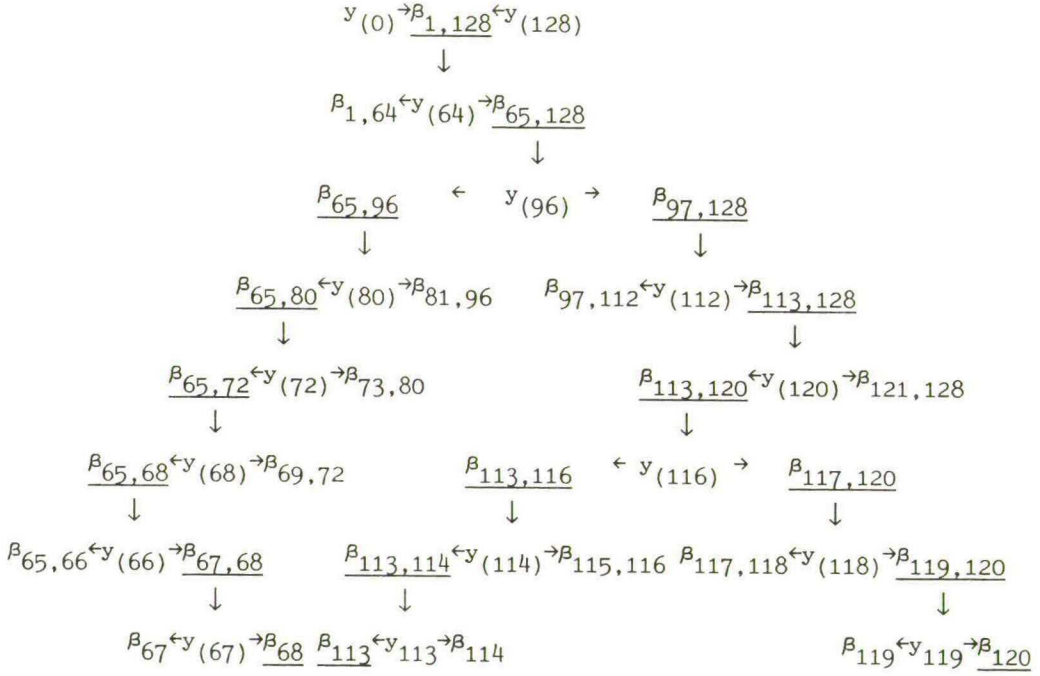


FIGURE 1. Example of Sequential Bifurcation.



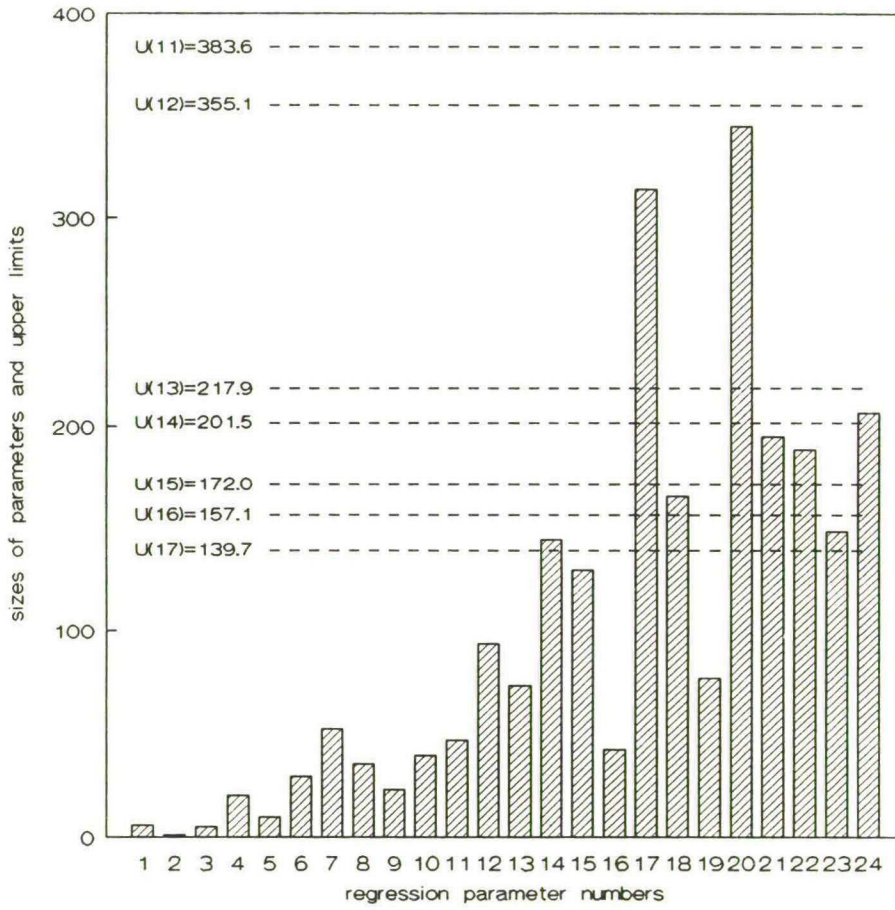


FIGURE 2. Main Effects and Upper Limits in Various Stages.

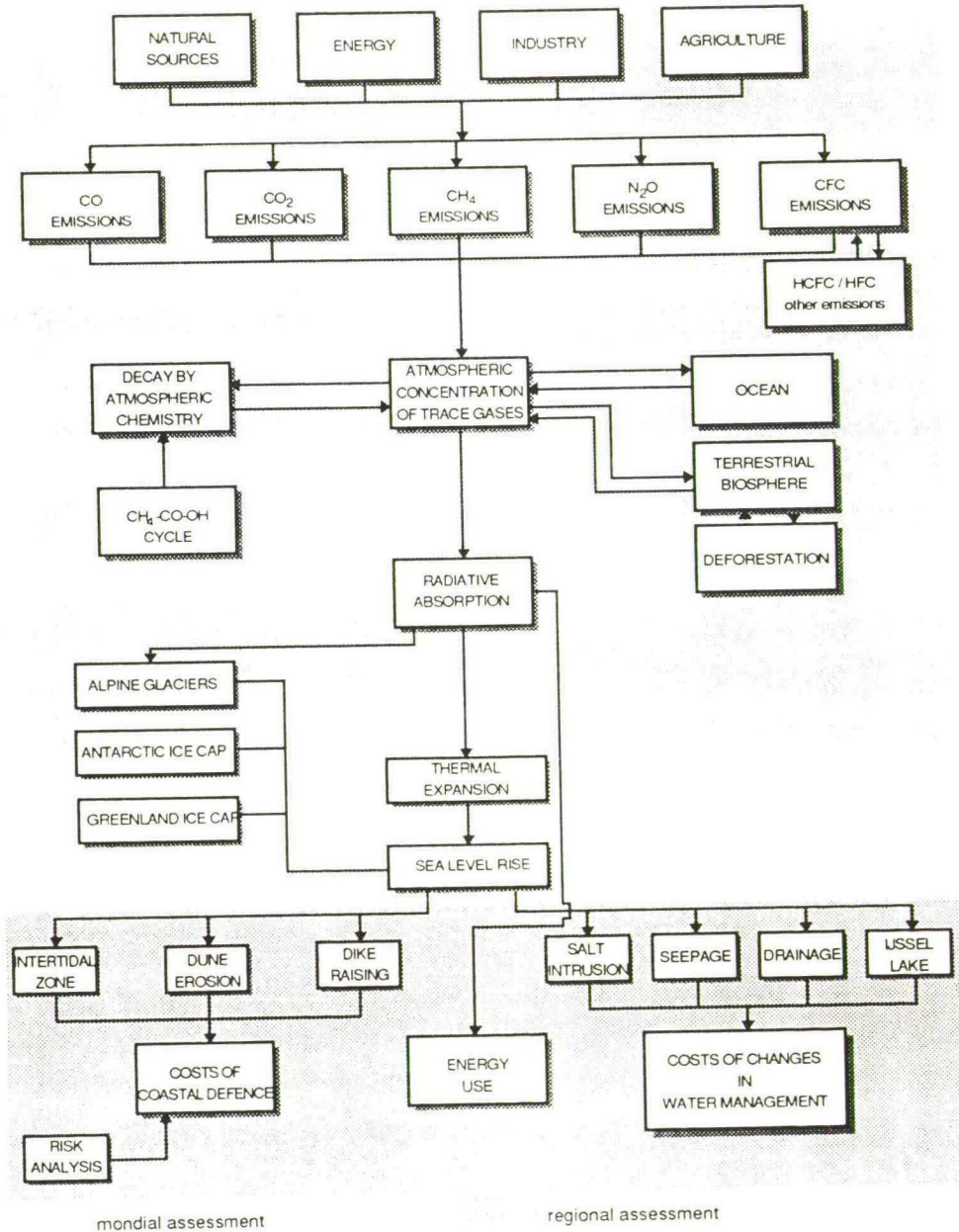


FIGURE 3. Integrated Model to Assess the Greenhouse Effect (IMAGE)

(Source: Rotmans, 1990, p.14)

## IN 1990 REEDS VERSCHENEN

- 419 Bertrand Melenberg, Rob Alessie  
A method to construct moments in the multi-good life cycle consumption model
- 420 J. Kriens  
On the differentiability of the set of efficient  $(\mu, \sigma^2)$  combinations in the Markowitz portfolio selection method
- 421 Steffen Jørgensen, Peter M. Kort  
Optimal dynamic investment policies under concave-convex adjustment costs
- 422 J.P.C. Blanc  
Cyclic polling systems: limited service versus Bernoulli schedules
- 423 M.H.C. Paardekooper  
Parallel normreducing transformations for the algebraic eigenvalue problem
- 424 Hans Gremmen  
On the political (ir)relevance of classical customs union theory
- 425 Ed Nijssen  
Marketingstrategie in Machtspectief
- 426 Jack P.C. Kleijnen  
Regression Metamodels for Simulation with Common Random Numbers: Comparison of Techniques
- 427 Harry H. Tigelaar  
The correlation structure of stationary bilinear processes
- 428 Drs. C.H. Veld en Drs. A.H.F. Verboven  
De waardering van aandelenwarrants en langlopende call-opties
- 429 Theo van de Klundert en Anton B. van Schaik  
Liquidity Constraints and the Keynesian Corridor
- 430 Gert Nieuwenhuis  
Central limit theorems for sequences with  $m(n)$ -dependent main part
- 431 Hans J. Gremmen  
Macro-Economic Implications of Profit Optimizing Investment Behaviour
- 432 J.M. Schumacher  
System-Theoretic Trends in Econometrics
- 433 Peter M. Kort, Paul M.J.J. van Loon, Mikuláš Luptacik  
Optimal Dynamic Environmental Policies of a Profit Maximizing Firm
- 434 Raymond Gradus  
Optimal Dynamic Profit Taxation: The Derivation of Feedback Stackelberg Equilibria

- 435 Jack P.C. Kleijnen  
Statistics and Deterministic Simulation Models: Why Not?
- 436 M.J.G. van Eijs, R.J.M. Heuts, J.P.C. Kleijnen  
Analysis and comparison of two strategies for multi-item inventory systems with joint replenishment costs
- 437 Jan A. Weststrate  
Waiting times in a two-queue model with exhaustive and Bernoulli service
- 438 Alfons Daems  
Typologie van non-profit organisaties
- 439 Drs. C.H. Veld en Drs. J. Grazell  
Motieven voor de uitgifte van converteerbare obligatieleningen en warrantobligatieleningen
- 440 Jack P.C. Kleijnen  
Sensitivity analysis of simulation experiments: regression analysis and statistical design
- 441 C.H. Veld en A.H.F. Verboven  
De waardering van conversierechten van Nederlandse converteerbare obligaties
- 442 Drs. C.H. Veld en Drs. P.J.W. Duffhues  
Verslaggevingsaspecten van aandelenwarrants
- 443 Jack P.C. Kleijnen and Ben Annink  
Vector computers, Monte Carlo simulation, and regression analysis: an introduction
- 444 Alfons Daems  
"Non-market failures": Imperfecties in de budgetsector
- 445 J.P.C. Blanc  
The power-series algorithm applied to cyclic polling systems
- 446 L.W.G. Strijbosch and R.M.J. Heuts  
Modelling (s,Q) inventory systems: parametric versus non-parametric approximations for the lead time demand distribution
- 447 Jack P.C. Kleijnen  
Supercomputers for Monte Carlo simulation: cross-validation versus Rao's test in multivariate regression
- 448 Jack P.C. Kleijnen, Greet van Ham and Jan Rotmans  
Techniques for sensitivity analysis of simulation models: a case study of the CO<sub>2</sub> greenhouse effect
- 449 Harrie A.A. Verbon and Marijn J.M. Verhoeven  
Decision-making on pension schemes: expectation-formation under demographic change



- 450 Drs. W. Reijnders en Drs. P. Verstappen  
Logistiek management marketinginstrument van de jaren negentig
- 451 Alfons J. Daems  
Budgeting the non-profit organization  
An agency theoretic approach
- 452 W.H. Haemers, D.G. Higman, S.A. Hobart  
Strongly regular graphs induced by polarities of symmetric designs
- 453 M.J.G. van Eijs  
Two notes on the joint replenishment problem under constant demand
- 454 B.B. van der Genugten  
Iterated WLS using residuals for improved efficiency in the linear model with completely unknown heteroskedasticity
- 455 F.A. van der Duyn Schouten and S.G. Vanneste  
Two Simple Control Policies for a Multicomponent Maintenance System
- 456 Geert J. Almekinders and Sylvester C.W. Eijffinger  
Objectives and effectiveness of foreign exchange market intervention  
A survey of the empirical literature
- 457 Saskia Oortwijn, Peter Borm, Hans Keiding and Stef Tijs  
Extensions of the  $\tau$ -value to NTU-games
- 458 Willem H. Haemers, Christopher Parker, Vera Pless and Vladimir D. Tonchev  
A design and a code invariant under the simple group  $Co_3$
- 459 J.P.C. Blanc  
Performance evaluation of polling systems by means of the power-series algorithm
- 460 Leo W.G. Strijbosch, Arno G.M. van Doorne, Willem J. Selen  
A simplified MOLP algorithm: The MOLP-S procedure
- 461 Arie Kapteyn and Aart de Zeeuw  
Changing incentives for economic research in The Netherlands
- 462 W. Spanjers  
Equilibrium with co-ordination and exchange institutions: A comment
- 463 Sylvester Eijffinger and Adrian van Rixtel  
The Japanese financial system and monetary policy: A descriptive review
- 464 Hans Kremers and Dolf Talman  
A new algorithm for the linear complementarity problem allowing for an arbitrary starting point
- 465 René van den Brink, Robert P. Gilles  
A social power index for hierarchically structured populations of economic agents

## IN 1991 REEDS VERSCHENEN

- 466 Prof.Dr. Th.C.M.J. van de Klundert - Prof.Dr. A.B.T.M. van Schaik  
Economische groei in Nederland in een internationaal perspectief
- 467 Dr. Sylvester C.W. Eijffinger  
The convergence of monetary policy - Germany and France as an example
- 468 E. Nijssen  
Strategisch gedrag, planning en prestatie. Een inductieve studie binnen de computerbranche
- 469 Anne van den Nouweland, Peter Borm, Guillermo Owen and Stef Tijs  
Cost allocation and communication
- 470 Drs. J. Grazell en Drs. C.H. Veld  
Motieven voor de uitgifte van converteerbare obligatieleningen en warrant-obligatieleningen: een agency-theoretische benadering
- 471 P.C. van Batenburg, J. Kriens, W.M. Lammerts van Bueren and R.H. Veenstra  
Audit Assurance Model and Bayesian Discovery Sampling
- 472 Marcel Kerkhofs  
Identification and Estimation of Household Production Models
- 473 Robert P. Gilles, Guillermo Owen, René van den Brink  
Games with Permission Structures: The Conjunctive Approach
- 474 Jack P.C. Kleijnen  
Sensitivity Analysis of Simulation Experiments: Tutorial on Regression Analysis and Statistical Design
- 475 An  $O(n \log n)$  algorithm for the two-machine flow shop problem with controllable machine speeds  
C.P.M. van Hoesel
- 476 Stephan G. Vanneste  
A Markov Model for Opportunity Maintenance
- 477 F.A. van der Duyn Schouten, M.J.G. van Eijs, R.M.J. Heuts  
Coordinated replenishment systems with discount opportunities
- 478 A. van den Nouweland, J. Potters, S. Tijs and J. Zarzuelo  
Cores and related solution concepts for multi-choice games
- 479 Drs. C.H. Veld  
Warrant pricing: a review of theoretical and empirical research
- 480 E. Nijssen  
De Miles and Snow-typologie: Een exploratieve studie in de meubelbranche
- 481 Harry G. Barkema  
Are managers indeed motivated by their bonuses?

- 482 Jacob C. Engwerda, André C.M. Ran, Arie L. Rijkeboer  
Necessary and sufficient conditions for the existence of a positive definite solution of the matrix equation  $X + A^T X^{-1} A = I$
- 483 Peter M. Kort  
A dynamic model of the firm with uncertain earnings and adjustment costs
- 484 Raymond H.J.M. Gradus, Peter M. Kort  
Optimal taxation on profit and pollution within a macroeconomic framework
- 485 René van den Brink, Robert P. Gilles  
Axiomatizations of the Conjunctive Permission Value for Games with Permission Structures
- 486 A.E. Brouwer & W.H. Haemers  
The Gewirtz graph - an exercise in the theory of graph spectra
- 487 Pim Adang, Bertrand Melenberg  
Intratemporal uncertainty in the multi-good life cycle consumption model: motivation and application
- 488 J.H.J. Roemen  
The long term elasticity of the milk supply with respect to the milk price in the Netherlands in the period 1969-1984
- 489 Herbert Hamers  
The Shapley-Entrance Game
- 490 Rezaul Kabir and Theo Vermaelen  
Insider trading restrictions and the stock market
- 491 Piet A. Verheyen  
The economic explanation of the jump of the co-state variable
- 492 Drs. F.L.J.W. Manders en Dr. J.A.C. de Haan  
De organisatorische aspecten bij systeemontwikkeling een beschouwing op besturing en verandering
- 493 Paul C. van Batenburg and J. Kriens  
Applications of statistical methods and techniques to auditing and accounting
- 494 Ruud T. Frambach  
The diffusion of innovations: the influence of supply-side factors
- 495 J.H.J. Roemen  
A decision rule for the (des)investments in the dairy cow stock
- 496 Hans Kremers and Dolf Talman  
An SLSPP-algorithm to compute an equilibrium in an economy with linear production technologies

- 497 L.W.G. Strijbosch and R.M.J. Heuts  
Investigating several alternatives for estimating the compound lead  
time demand in an  $(s,Q)$  inventory model



Bibliotheek K. U. Brabant



17 000 01066386 3